

6/3/98 rev. 10/20/98, 3/23/99, 6/2/99, 11/15/2001

## II. StEPS General Item Imputation Methods

### 1. Parameters for Imputation Methods

#### Parameters

<i>Name</i>	<i>Description</i>	<i>Value</i>	<i>Methods Used In</i>
NUMAUX	<b>Number of Specified Auxiliary Variables.</b> The number of auxiliary variables to be used with a method.	Non-negative integer.	SUM                      ATREND RESIDUA                AUXRAT PRODUCT                VALUE MEAN                    SIMPREG RATIO                    MULTREG Custom method
AUX1- AUX10	<b>Auxiliary Variables.</b> Auxiliary variables for item imputation methods.	Item name, control variable name, recoded variable name, or numeric constant. Item names and recoded variable names must include the two digit relative stat period and, if applicable, the one digit data prefix, E=edited or A=adjusted.  ✂No entry= indicated by a blank. Default is a blank.	SUM                      ATREND RESIDUA                AUXRAT PRODUCT                VALUE MEAN                    SIMPREG RATIO                    MULTREG Custom method
EXTAUX1- EXTAUX10	<b>Extra Operations Applied to Corresponding Auxiliary Variables.</b> Sometimes an auxiliary variable needs to be multiplied by a factor or some other operation. EXTAUX <sub>j</sub> contains the SAS code for this.  The value of AUX <sub>j</sub> concatenated with	Must start with a SAS operation followed by:  Item name, control variable name, recoded variable name, or numeric constant. Item names and recoded variable names must include the two digit relative stat period and, if applicable, the one digit data prefix, E=edited or A=adjusted.	SUM                      ATREND RESIDUA                AUXRAT PRODUCT                VALUE MEAN                    SIMPREG RATIO                    MULTREG Custom method

### Parameters

<i>Name</i>	<i>Description</i>	<i>Value</i>	<i>Methods Used In</i>
	the value of EXTAUX <sub>j</sub> must form a valid SAS expression.	»No entry= indicated by a blank field. Default is a blank.	
NUMREG	<b>Number of Specified Regression Coefficients.</b> The number of regression coefficients to be used with a regression method.	Non-negative integer. Default is a blank.	SIMPREG MULTREG Custom method
REG1– REG10	Regression coefficients	Item name, control variable name, recoded variable name, or numeric constant. Item names and recoded variable names must include the two digit relative stat period and, if applicable, the one digit data prefix, E=edited or A=adjusted.  »No entry= indicated by a blank. Default is a blank.	SIMPREG MULTREG Custom method
CAT	A categorical item. The categorical item can be a recode.	Item name, control variable name, or recoded variable name. Item names and recoded variable names must include the two digit relative stat period and, if applicable, the one digit data prefix, E=edited or A=adjusted.  »No entry= indicated by a blank. No category variable is indicated by »_NULL_«, that is, imputation base calculated from all eligible cases. Default is _NULL_.	MEAN RATIO  AUXRAT Custom method
U	<b>Upper bound.</b> Upper bound for a data value to be included in the computation of a mean or a ratio of data	Numerical constant.  »No entry= indicated by a blank. »No upper bound=	MEAN RATIO  AUXRAT Custom method

### Parameters

<i>Name</i>	<i>Description</i>	<i>Value</i>	<i>Methods Used In</i>
	values to be included in the computation of a ratio-of-identicals.	indicated by >_NULL_=	
L	<b>Lower bound.</b> Lower bound for a data value to be included in the computation of a mean or a ratio of data values to be included in the computation of a ratio-of-identicals.	Numeric constant.  ⌘No entry= indicated by a blank. ⌘No lower bound= indicated by >_NULL_=	MEAN RATIO  AUXRAT Custom method
MINNUM	The minimum number of cases needed to calculate a mean or a ratio-of-identicals.	Non-negative integer.  ⌘No entry= indicated by a blank. The default value is 10	MEAN RATIO  AUXRAT Custom method

### Parameters for Custom Methods

<i>Name</i>	<i>Description</i>	<i>Value</i>	<i>Methods Used In</i>
NUMPAR	Number of additional parameters	A non-negative integer.	Custom method
PARNM1– PARNM20	Name of additional parameter	Name of a custom item imputation method other than NUMAUX, AUX <sub>n</sub> , NUMREG, REG <sub>n</sub> , CAT, MINNUM, U or L.  ⌘No entry= indicated by a blank.	Custom method
PARVL1–	Value of additional parameter	Item name, control variable name, recoded variable name, or numeric	Custom method

### Parameters for Custom Methods

<i>Name</i>	<i>Description</i>	<i>Value</i>	<i>Methods Used In</i>
PARVL20		constant. Names must include the two digit relative stat period and, if applicable, the one digit data prefix, E=edited or A=adjusted.  No entry=indicated by a blank. Default is a blank.	

## 2. Imputation Methods

This section describes the imputation methods available for general imputation.

### NOTATION

$v$  = the current statistical period value of the item (OBJECT) inputted to the imputation method.

$v'$  = the value for the item (OBJECT) outputted from the imputation method.

$z_j$  = the value of an auxiliary item<sup>1</sup> (AUX $j$  or AUX $j$  concatenated with EXTAUX $j$ ) for  $j = 1, \dots, 10$ .

$S(f)$  = the sum of item  $f$ , the sum may be simple or weighted.

$(S(f_1)/S(f_2))_1$  = the Ratio of Identicals of items  $f_1$  to  $f_2$ . This is the ratio of two sums both of which use all eligible records that have reported both items.

---

<sup>1</sup> Examples. 1. EXTAUX1 is blank. If AUX1 = ~~ECSAL00~~ and the value of ECSAL00 = 196 then the value of  $z_1$  is 196.  
2. EXTAUX1 is ~~>/ EPSAL00~~ and the value of EPSAL00 is 224. AUX1 is the same as above. AUX1 concatenated with EXTAUX1 is ~~>ECSAL00 / EPSAL00~~ and the value of  $z_1$  is  $196/224 = 0.875$ .

$b_{v,z_j}$  = the regression coefficient (REGj) for data item  $v$  and auxiliary item  $z_j$  for  $j = 1, \dots, 10$ .

Many of the imputation methods require the computation of a ratio-of-identicals or a mean. They use the parameters U and L.

U = the upper bound for an item ratio,  $f_1 / f_2$ , used in a ratio-of-identicals imputation or the upper bound of an item,  $f$ , used in MEAN imputation. The default is plus infinity.

L = the lower bound for an item ratio,  $f_1 / f_2$ , used in a ratio-of-identicals imputation or the upper bound of an item,  $f$ , used in MEAN imputation. The default is minus infinity.

Items ineligible for the computation of a ratio-of-identicals satisfy at least one of the following conditions:

- N Items where the entire ID is to be bypassed for the imputation base (BYIMPB=B).
- N Items where IMPACT is N, Y, or X.
- N Items in the *Imputation Reject* file.
- N Items that are missing.
- N IDs that do not meet user's include clause, i.e., do not satisfy CRE8IMPB if CRE8IMPB is not blank.
- N Numerator items ( $f_1$ ) in ratio-of-identicals in which
  - 1) the corresponding denominator item ( $f_2$ ) is zero, missing or on the Imputation Reject file,  
or
  - 2) the ratio  $f_1/f_2$  does not fall within user specified (or default) upper (U) and lower (L) bounds.
- N Denominator items ( $f_2$ ) in a ratio-of-identicals in which
  - 1)  $f_2=0$ ,  
or
  - 2) the corresponding numerator item ( $f_1$ ) is missing or on the Imputation Reject file,  
or
  - 3) the ratio  $f_1/f_2$  does not fall within user specified (or default) upper (U) and lower (L) bounds.

- N Denominator and numerator items ( $f_1, f_2$ ) in a ratio-of-identicals when the numerator or denominator has a data flag set as not reported (data flag of F, I, M, V or T).

Items ineligible for the computation of a mean satisfy at least one of these following conditions:

- N Items where the entire ID is to be bypassed (BYIMPB=B).
- N Items where IMPACT is N, Y, or X.
- N Items on the *Imputation Reject* file.
- N Items that are missing.
- N IDs that do not satisfy CRE8IMPB if CRE8IMPB is not blank.
- N Items that do not fall within user specified (or default) upper (U) and lower (L) bounds.

<b>Logical and Direct Substitution Imputation Methods</b>					
<i>Name</i>	<i>Definition</i>	<i>ImpFlg</i>	<i>Implied Conditions</i>	<i>Parameters</i> <sup>2</sup>	<i>Specification</i>
SUM	Imputes the sum of auxiliary variables.	A	There are auxiliary parameters 1 through NUMAUX. The values of the auxiliary parameters are non-missing.	NUMAUX AUX1( $z_1$ ), EXTAUX1, ..., AUX $n$ ( $z_n$ ) EXTAUX $n$	$v' = z_1 + z_2 + \dots + z_n$
RESIDUA	Imputes an auxiliary variable minus the sum of detail auxiliary variables.	N	There are auxiliary parameters 1 through NUMAUX. The values of the auxiliary parameters are non-missing.	NUMAUX AUX1( $z_1$ ), EXTAUX1, ..., AUX $n$ ( $z_n$ ) EXTAUX $n$	If $z_1 - (z_2 + \dots + z_n) \geq 0$ or NEGFLG of $v$ is >Y= then $v' = z_1 - (z_2 + \dots + z_n)$
PRODUCT	Imputes the product of two auxiliary variables.	X	NUMAUX=2. The values of AUX1 and AUX2 are non-missing.	NUMAUX AUX1( $z_1$ ), EXTAUX1, AUX2( $z_2$ ), EXTAUX2	$v' = z_1 \times z_2$
VALUE	Imputes the value of the auxiliary variable.	V	AUX item is non-missing.	NUMAUX AUX1( $z_1$ )	$v' = z_1$

<sup>2</sup> $z_j$  is AUX $j$  or AUX $j$  concatenated with EXTAUX $j$ .

				EXTAUX1	
--	--	--	--	---------	--

<b>Mean Imputation</b>					
<i>Name</i>	<i>Definition</i>	<i>ImpFlg</i>	<i>Implied Conditions</i>	<i>Parameters</i> <sup>3</sup>	<i>Specification</i>
MEAN	Imputes the mean value of an auxiliary variable using all eligible records. The auxiliary can be the item from the same or a different statistical period.	M	$\bar{z}_I$ is non-missing. U>L.	NUMAUX AUX1(z <sub>1</sub> ) EXTAUX1 U L CAT MINNUM	$v' = \bar{z}_I$

<b>Ratio Imputation</b>					
<i>Name</i>	<i>Definition</i>	<i>ImpFlg</i>	<i>Implied Conditions</i>	<i>Parameters</i> <sup>3</sup>	<i>Specification</i>

---

<sup>3</sup>z<sub>j</sub> is AUX<sub>j</sub> or AUX<sub>j</sub> concatenated with EXTAUX<sub>j</sub>.



<b>Ratio Imputation</b>					
<i>Name</i>	<i>Definition</i>	<i>ImpFlg</i>	<i>Implied Conditions</i>	<i>Parameters</i> <sup>3</sup>	<i>Specification</i>
RATIO	Imputes the value from an auxiliary variable adjusted by the corresponding ratio-of-identicals. The auxiliary can be the item from a different statistical period.	I	The ratio-of-identicals and the value of AUX1 are non-missing. U>L.	NUMAUX AUX1(z <sub>1</sub> ) EXTAUX1 U L CAT MINNUM	$v' = (S(v) / S(z_1))_1 z_1$
ATREND	Imputes the value of the AUX1 adjusted by the ratio of two other auxiliary items.	T	There are three auxiliary items. NUMAUX=3. The values of the auxiliary items are non-missing. AUX3 > 0.	NUMAUX AUX1(z <sub>1</sub> ) AUX2(z <sub>2</sub> ) AUX3(z <sub>3</sub> ) EXTAUX1 EXTAUX2 EXTAUX3	$v' = z_1 (z_2 / z_3)$  Often, z <sub>1</sub> is v from a different statistical period. z <sub>2</sub> is an auxiliary variable from the same statistical period as v and z <sub>3</sub> is the same auxiliary variable as z <sub>2</sub> but from the same statistical period as z <sub>1</sub> .
AUXRAT	Imputes the value of AUX1 adjusted by the ratio-of-identicals for AUX2 and AUX3.	J	The ratio-of-identicals and the value of AUX1 are non-missing. U>L. There are three auxiliary items. NUMAUX=3.	NUMAUX AUX1(z <sub>1</sub> ) AUX2(z <sub>2</sub> ) AUX3(z <sub>3</sub> ) EXTAUX1 EXTAUX2 EXTAUX3 U L	$v' = z_1 (S(z_2) / S(z_3))_1$  Often, z <sub>1</sub> is v from a different statistical period. z <sub>2</sub> is an auxiliary variable from the same statistical period as v and z <sub>3</sub> is the same auxiliary variable as z <sub>2</sub> but from the same statistical

<b>Ratio Imputation</b>					
<i>Name</i>	<i>Definition</i>	<i>ImpFlg</i>	<i>Implied Conditions</i>	<i>Parameters</i> <sup>3</sup>	<i>Specification</i>
				CAT MINNUM	period as $z_1$ .

<b>Regression Imputation</b>					
<i>Name</i>	<i>Definition</i>	<i>ImpFlg</i>	<i>Implied Conditions</i>	<i>Parameters</i> <sup>4</sup>	<i>Specification</i>
SIMPREG	Imputes an auxiliary variable times a regression coefficient.	G	Auxiliary item is non-missing. The regression coefficient is non-missing.	NUMAUX AUX1( $z_1$ ) EXTAUX1 NUMREG REG1( $\mathbf{b}_{v,z_1}$ )	$\mathbf{v}' = \mathbf{b}_{v,z_1} z_1$
MULTREG	Imputes the sum of auxiliary variables time regression coefficients.	H	Auxiliary items are non-missing. There are auxiliary variables 1 through NUMAUX. The regression coefficients are non-missing. There are	NUMAUX AUX1( $z_1$ ), EXTAUX1, ..., AUX $n$ ( $z_n$ ) EXTAUX $n$	$\mathbf{v}' = \mathbf{b}_{v,z_1} z_1 + \dots + \mathbf{b}_{v,z_n} z_n$

---

<sup>4</sup> $z_j$  is AUX $j$  or AUX $j$  concatenated with EXTAUX $j$ .

			regression coefficients 1 through NUMAUX.	NUMREG REG1( $\mathbf{b}_{v,z1}$ )  , ..., REGn( $\mathbf{b}_{v,zn}$ )	
--	--	--	--	--	--

## **Item Imputation Source Data Flag**

The Item Imputation Source Data Flag (the e2 flag) for an item imputation is the data flag for the source of the item imputation when the methods VALUE, RATIO, ATREND, AUXRAT, and SIMPREG have successfully imputed the item. For these methods, the source for the imputation is the AUX1 variable. Item imputation assigns the value of the Data Flag for the AUX1 variable to the e2 flag for the item. The default value for the e2 flag is blank.

The e2 flag will be set to blank when any of the other item imputation methods successfully imputes. The e2 flag will not be changed when an item imputation method does not successfully impute.

## **Example of the Use of AUXRAT for Percentage Imputation**

The Annual Survey of Communication Services<sup>BR</sup>Radio and Television Broadcasting collects data on the percentage of a firm's total operating revenue from radio stations and television stations. These two items have key codes 716 and 717 respectively. The key code for total operating revenue is 002. The formula for estimating a percentage by ratio imputation is  $(S(\text{percentages} \times \text{base}) / S(\text{base}))_1$ . Use AUXRAT to impute these two percentages. The figure below shows the >Add AUXRAT Item Imputation= screen for estimating the percentage for key code 716.

Disclosure Prohibited - Title 13 U.S. Code

Survey: [SITS\\_I](#) Stat Period: [1997A1](#) Date: [16OCT98](#)

Spec Number:  Object:  Method:  Spec Status: ☒ Active ☐ Inactive ☐ Pending

Source

Data Items:   
  
  
  
  
  
  
  
  
  
 Tot:

Control Items:   
  
  
  
  
  
  
  
  
  
 Tot:

Imputation Condition:

Target

Aux	Extra
1	
E71600	* E00200
E00200	

Tot:

Cat:  Minimum:   
 L:  U:

Other:

Description:

Last Update:  By:

AUXRAT imputes by multiplying the value of the first auxiliary item,  $AUX1(z_1)$ , by a ratio-of-identicals of the other two auxiliary items,  $AUX2(z_2)$  and  $AUX3(z_3)$ . The estimate for the percentage is just the ratio-of-identicals so that  $AUX1$  must be set to one. A constant is an acceptable value for an auxiliary variable. The base for the ratio-of-identicals is total operating revenue, key code 002.  $AUX3$  is the base so that  $AUX3$  is  $\$E00200$ .  $z_2$  is the numerator for the ratio-of-identicals.  $z_2$  can be either  $AUX2$  or  $AUX2$  concatenated with  $EXTAUX2$ . In this case, you need both  $AUX2$  and  $EXTAUX2$ . The screen above shows that  $AUX2$  is  $\$E71600$  and  $EXTAUX2$  is  $\$* E00200$ .  $AUX2 = \$E00200$  and  $EXTAUX2 = \$* E71600$  will also be acceptable. The above screen shows the values of the auxiliary variables and the  $EXTAUX$  variables as entered into the screen. Imputation using AUXRAT for operating revenue from television stations, key code 717, is set up the same as for key code 716 substituting  $\$E71700$  for  $\$E71600$ . At the end of each imputation step, the imputed percentage is rounded to an integer.

Once the percentages have been imputed, the sum of the reported and imputed values may not add to 100. Use balance complex imputation (methods RAKE or RAKEIMP) to bring the sum to 100.

## CAT and MINNUM

CAT and MINNUM are parameters used with the mean and ratio-of-identicals methods. The value of CAT can be a categorical item with discrete values such as SIC or a recode derived from other categorical items. There needs to be a recode specification (Specification type=R) for a recode to be used as the value of CAT.

The parameter CAT specifies a categorical item for which the ratio-of-identicals or mean must be calculated for each distinct value of this item. These ratios and means will be placed into the imputation data base. Missing values are acceptable for the ratios-of-identicals and means.

A survey may want to have successive imputations using the same method but using broader categories until that item has been imputed. To do this, separate specifications for an item using each of the different categorical variables must be specified. In each of these specifications, the value for CAT will be a categorical item or a recode. Use method order to specify the sequence in which imputation will be tried with these categorical items. This is usually from most detailed (smallest method order) to the least detailed (largest method order) for an object. The categorical items do not need to be nested but usually are. For example, SIC4 is nested within SIC3 and SIC3 is nested within SIC2. SIC3 and SIC2 could be recodes of SIC4.

MINNUM specifies the minimum number of cases required to calculate the ratio-of-identicals or the mean. If the number of eligible cases available to calculate the ratio-of-identicals or mean in a category is fewer than MINNUM, then the specification will not be used to impute items using that specific category. No imputation will be done when the value of one of the parameters is missing.